Hybrid Retinal Image Registration

Thitiporn Chanwimaluang, *Student Member, IEEE*, Guoliang Fan, *Senior Member, IEEE*, and Stephen R. Fransen

***Abstract*—This work studies retinal image registration in the context of the National Institutes of Health (NIH) Early Treatment Diabetic Retinopathy Study (ETDRS) standard. The ETDRS imaging protocol specifies seven fields of each retina and presents three major challenges for the image registration task. First, small overlaps between adjacent fields lead to inadequate landmark points for feature-based methods. Second, the non-uniform contrast/intensity distributions due to imperfect data acquisition will deteriorate the performance of area-based techniques. Third, high-resolution images contain large homogeneous nonvascular/texureless regions that weaken the capabilities of both feature-based and area-based techniques. In this work, we propose a hybrid retinal image registration approach for ETDRS images that effectively combines both area-based and feature-based methods. Four major steps are involved. First, the vascular tree is extracted by using an efficient local entropy-based thresholding technique. Next, zeroth-order translation is estimated by maximizing mutual information based on the binary image pair (area-based). Then image quality assessment regarding the ETDRS field definition is performed based on the translation model. If the image pair is accepted, higher-order transformations will be involved. Specifically, we use two types of features, landmark points and sampling points, for affine/quadratic model estimation. Three empirical conditions are derived experimentally to control the algorithm progress, so that we can achieve the lowest registration error and the highest success rate. Simulation results on 504 pairs of ETDRS images show the effectiveness and robustness of the proposed algorithm.**

***Index Terms*—Area-based registration, feature-based registration, mutual information (MI), retinal image registration, vascular tree extraction.**

I. Introduction

IMAGE registration is a fundamental problem to several image processing and computer vision applications [1], [2]. In the case of medical imaging, disease diagnosis and treatment planning are often supported by multiple images acquired from the same patient. Image registration techniques, hence, are needed in order to integrate the information gained from several images to obtain a comprehensive understanding [3]–[5]. A broad range of image registration methods have been proposed for different medical imaging applications including retinal image registration. Various criteria, e.g., modalities, dimensionalities, elasticity of the transformation, have been proposed to categorize registration methods [1]–[4]. Typically, retinal image registration techniques are classified as feature-based and area-based methods.

Manuscript received December 21, 2004; revised June 14, 2005. This work was supported by an OHRS award of Project Number HR03-33 from the Oklahoma Center for the Advancement of Science and Technology (OCAST).

T. Chanwimaluang and G. Fan are with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078 USA (e-mail: thitipo@okstate.edu; glfan@okstate.edu).

S. R. Fransen is with the Inoveon Corporation, Department of Ophthalmology, University of Oklahoma Health Sciences Center; Dean A. Mcgee Eye Institute, Oklahoma City, OK 73104 USA (e-mail: stephen-fransen@inoveon.com).

Digital Object Identifier 10.1109/TITB.2005.856859

Area-based techniques are generally based on pixel intensities and certain optimized objective functions, such as least mean square error, cross-correlation, phase correlation, or mutual information, [6]–[12]. In the case of retinal image registration, area-based approaches are often used in multimodal or temporal image registration applications. In [8], mutual information is used as a similarity measure and simulated annealing is employed as a searching technique. In [7], the measure of match (MOM) is proposed as an objective function and the genetic algorithm is chosen to be the optimization technique. Nevertheless, the searching space of transformation models (affine, bilinear, and projective) is huge. The greater the geometric distortion between the image pair, the more complicated the searching space. Typically, there are two major factors that may degrad the performance of area-based methods: nonconsistent/nonuniform contrast within an image and large homogeneous/textureless areas.

Feature-based methods are somewhat similar to manual registration [13]–[26]. The approach assumes that point correspondences are available in both images, and the registration process is performed by maximizing a similarity measure computed from the correspondences. In [15], the bifurcation points of a vascular tree, also called landmark points, are labeled with surrounding vessel orientations. An angle-based invariant is then computed to give a probability for every two matching points. After that, the Bayesian–Hough transform is used to sort the transformations according to their respective likelihoods. In [13], the similarity matrix for all possible correspondences is computed based on the orientations of vascular centerlines and the similarity measure is converted to a prior probability. The transformation is estimated in a hierarchical way, from the zeroth-order model to the first-order model and finally to the second-order model. Nonetheless, sufficient feature points have to be available. In [18], the dual-bootstrap iterative closest point (dual-bootstrap ICP) algorithm was introduced. The approach starts from one or more initial, low-order estimates that are only accurate in small image regions called bootstrap regions. In each bootstrap region, the method iteratively refines the transformation estimation, expands the bootstrap region, and tests to see if a higherorder model can be used. The method requires accurate initialization of at least one point correspondence. High success rates were reported in [18]. The performance of feature-based methods largely depends on sufficient and/or reliable correspondences, especially, when the overlapping part of an image pair is very limited or when there are mis-matched correspondences.

In this paper, we study retinal image registration in the context of the National Institutes of Health (NIH), Early Treatment Diabetic Retinopathy Study (ETDRS) standard protocol [27]. The ETDRS protocol defines seven 30° fields of each retina with specific field coverage. A robust ETDRS image

1089-7771/$20.00 © 2006 IEEE

registration algorithm is required to: 1) assess image quality in terms of ETDRS field coverage and 2) support ETDRS-based disease staging. Three major challenges are present. First, small overlaps between adjacent fields lead to inadequate landmark points (crossovers and bifurcations) for feature-based methods. Second, the contrast and intensity distributions within an image are not spatially uniform or consistent. This can deteriorate the performance of area-based techniques. Third, high-resolution ETDRS images contain large homogeneous nonvascular/textureless regions which result in difficulties for both feature-based and area-based techniques. In this work, we propose an ETDRS image registration algorithm which effectively combines both area-based and feature-based methods. There are three major steps in the proposed algorithm. First, binary vascular trees are extracted from retinal images using efficient local entropy-based thresholding method [28], [29]. Next, zeroth-order translation is estimated by maximizing mutual information based on the binary image pair (area-based). Specifically, a local entropy-based peak selection scheme and a multi-resolution searching strategy are developed to improve the accuracy and efficiency of translation estimation. Third, we use two types of features, landmark points and sampling points, for higher order transformation estimation. Sampling points, which are acquired by imposing a grid onto the thinned vascular tree, are only introduced when landmark points do not meet certain criteria. Simulation results on 504 pairs of ETDRS retinal images show the effectiveness and robustness of the proposed registration algorithm.

II. Preliminaries

A. NIH ETDRS Protocol

According to the statistics from the American Diabetes Association (http://www.diabetes.org), 18.2 million people in the United States, or 6.3% of the population, have diabetes. While an estimated 13 million have been diagnosed with diabetes, unfortunately, one-third of these people are unaware that they have this disease. Diabetes is the leading cause of blindness among working-age Americans, and many patients with vision-threatening diabetic retinopathy remain asymptomatic until blindness occurs. The great majority of this blindness can be prevented with proper eye examination and treatment by ophthalmologists who rely on the results of randomized clinical trials to guide their treatment of patients with diabetes.

NIH's ETDRS is a multicenter, randomized clinical trial designed to evaluate treatment of patients with nonproliferative or early proliferative diabetic retinopathy. A total of 3711 patients were recruited to be followed for a minimum of 4 years to provide long-term information on the risks and benefits of the treatments under study. The study demonstrated a statistically significant reduction in severe visual loss for those eyes with early treatment [27]. The ETDRS protocol standardizes retinal imaging and lesion classification, it also develops an internationally recognized disease severity scale for diabetic retinopathy. ETDRS protocols have become the "gold standard" for evaluating diabetic retinopathy [30] and diabetic macular edema [31].

TABLE I
Field Coverage Specification

Fields	Specifications
Field 1	Centered at optic disc.
Field 2	Centered at macula.
Field 3	The center of the macula appears approximately mid-way between the edge and the center of the field.
Field 4	The lower edge of the field is tangent to a horizontal line passing through the upper edge of the optic disc & the nasal edge of the field is tangent to a vertical line passing through the center of the disc.
Field 5	The upper edge of the field is tangent to a horizontal line passing through the lower edge of the optic disc & the nasal edge of the field is tangent to a vertical line passing through the center of the disc.
Field 6	The lower edge of the field is tangent to a horizontal line passing through the upper edge of the optic disc & the temporal edge of the field is tangent to a vertical line passing through the center of the disc.
Field 7	The upper edge of the field is tangent to a horizontal line passing through the lower edge of the optic disc & the temporal edge of the field is tangent to a vertical line passing through the center of the disc.

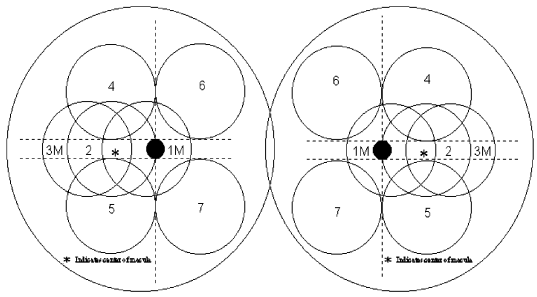


Fig. 1. ETDRS seven-standard fields (right and left eyes).

B. Image Quality Assessment (IQA): Field Definition

The ETDRS protocol specifies seven stereoscopic 30° fields of each eye, as defined in Table I and Fig. 1. The overlap of field pairs 1 and 2 (or fields $1/2$)[1] as well as that of fields $2/3$ are roughly 50% of the image size. For other field pairs, the overlaps are typically less than 25%. It is worth mentioning that the field displacements are not always consistent and depend on patient cooperation and photographer's skills. The importance of the ETDRS protocol and the challenges in its practical implementation call for automated software tools for image quality assessment (IQA) that checks the relative positions, i.e., horizontal/vertical displacements, of every image pair according to Table I. By comparing the offset, i.e., T_o, which is the difference between the desired vertical/horizontal displacements and actual ones, with the diameter of optic disc (DD), an image pair is categorized as good $(T_o < 1/2\text{DD})$, fair $(1/2\text{DD} \leq T_o \leq 1\text{DD})$, or poor $(T_o > 1\text{DD})$. Therefore, the IQA of ETDRS field definition boils down to a problem of image registration followed by displacement verification. We will briefly review some technical background of retinal image registration in the following.

[1]The notation of "fields $1/2$" indicates that field 1 is the fixed image (the model image) and field 2 is the image being mapped to (the distorted image)

C. Global and Local Entropy

The global entropy, or the entropy, of a n-state system is defined as a function of the state probability [32]

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{1}$$

where p_i is the probability of state i. In the case of an image, the entropy is not an adequate measure since image pixel intensities are not independent of each other. Different images with identical histograms will always yield the same value of entropy since the spatial distribution is not taken into account in the global entropy computation. The local entropy [33], on the other hand, can better distinguish two images in terms of their spatial structures, because it considers the dependency of image pixel intensities. The local entropy is defined as

$$H^{(2)} = -\sum_{i}\sum_{j} p_{ij} \log_2 p_{ij} \tag{2}$$

where p_{ij} is the probability of cooccurrence of the intensity levels, i and j. The cooccurrence matrix is an $L \times L$ dimensional matrix (L is the number of intensity levels) [33]. It indicates the transition of pixel intensities between adjacent pixels. The local entropy, therefore, can indicate the spatial structure/pattern of an image. The local entropy or second-order entropy will be used in this research for two main purposes: 1) selecting the threshold for vascular tree extraction and 2) choosing an optimal match out of multiple competitive matches in the translation estimation.

D. Area-Based Retinal Image Registration

As mentioned before, area-based retinal image registration techniques are usually based on image pixel intensities and certain optimization functions. Specifically, we focus on mutual information (MI) that shows the similarity between one image pair based on the histograms and the joint histogram. The definition of mutual information can be presented in various ways [34]. Here, we use the definition as follows:

$$\mathrm{MI}(\mathbf{I}_u, \mathbf{I}_v) = H(\mathbf{I}_u) + H(\mathbf{I}_v) - H(\mathbf{I}_u, \mathbf{I}_v) \tag{3}$$

where $H(\mathbf{I}_u)$ is the global entropy, defined in (1), of image $\mathbf{I}_u$. $H(\mathbf{I}_u, \mathbf{I}_v)$ is the joint entropy, i.e., the entropy of the joint probability distribution of images $\mathbf{I}_u$ and $\mathbf{I}_v$.

For the registration purpose, the MI is computed for the overlap of two images. It is, therefore, sensitive to the size of overlaps. According to [34], the overlap size influences the mutual information measure in two ways. First, a decrease in the overlap size decreases the number of samples, which degrades the estimation of statistical probability distributions. Second, it has been shown in [35] that the MI computation is not robust when the overlaps are too small. The entropy correlation coefficient (ECC) is a normalized measure of MI [36], [37], which is less sensitive to area changes in the overlaps, as defined in the following:

$$\mathrm{ECC}(\mathbf{I}_u, \mathbf{I}_v) = 2 - \frac{2H(\mathbf{I}_u, \mathbf{I}_v)}{H(\mathbf{I}_u) + H(\mathbf{I}_v)} \tag{4}$$

TABLE II
TRANSFORMATION MODELS

Model	Transformation Models	Degree of Freedom
Translation	$\begin{pmatrix} p_x \\ p_y \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & m_2 \\ 0 & 1 & m_8 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_x \\ q_y \\ 1 \end{pmatrix}$	2
Affine	$\begin{pmatrix} p_x \\ p_y \\ 1 \end{pmatrix} = \begin{pmatrix} m_0 & m_1 & m_2 \\ m_6 & m_7 & m_8 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q_x \\ q_y \\ 1 \end{pmatrix}$	6
Quadratic	$\begin{pmatrix} p_x \\ p_y \end{pmatrix} = \begin{pmatrix} m_0 & m_1 & m_2 & m_3 & m_4 & m_5 \\ m_6 & m_7 & m_8 & m_9 & m_{10} & m_{11} \end{pmatrix} \begin{pmatrix} q_x \\ q_y \\ 1 \\ q_x^2 \\ q_y^2 \\ q_x q_y \end{pmatrix}$	12

where $H(\mathbf{I}_u)$ and $H(\mathbf{I}_v)$ represent global entropy of images $\mathbf{I}_u$ and $\mathbf{I}_v$, respectively. $H(\mathbf{I}_u, \mathbf{I}_v)$ is the joint entropy between two images. It is shown that the ECC is generally a better option for registering images since it is less susceptible to different sizes of overlaps [36], [37].

E. Feature-Based Retinal Image Registration

Feature-based methods relies on correspondence points in both images. The matching process identifies reliable correspondences by maximizing an objective function related to features. Then the transformation is estimated by minimizing correspondences' displacement, i.e., the registration error

$$\hat{\mathbf{M}} = arg \min_{\mathbf{M}} median_{\mathbf{p}\in\mathbf{P}} \min_{\mathbf{q}\in\mathbf{Q}} \|\mathbf{p} - T(\mathbf{q}; \mathbf{M})\|^2 \tag{5}$$

where $\mathbf{P}$ and $\mathbf{Q}$ denote the feature point sets from two images. $T(\mathbf{q}; \mathbf{M})$ represents the transformation operation of point $\mathbf{q}$ given model $\mathbf{M}$. The *median* function can also be used in (5) that is less sensitive to outliers compared with the *mean* function [13], [14]. As listed in Table II, the transformation models often used include the translation model, the affine model, and the quadratic model [8], [13], [14], [18]. The translation model consists of two parameters characterizing the displacements along horizontal and vertical directions. The affine model describes translation, rotation, shearing, and scaling. However, the affine model cannot address the nonlinearity, i.e., retina's curvature. The quadratic transformation has 12 parameters and it can cope with nonlinearity. The estimated transformation model $\hat{\mathbf{M}}$ can be further adjusted by using the Iterative Closest Point algorithm (ICP) to refine correspondences [38]. The ICP is a procedure for iteratively matching a set of points in two images. Given an initial transformation model, for $\mathbf{p} \in \mathbf{P}$, we need to find the closest point $\mathbf{q} \in \mathbf{Q}$ by following:

$$d(\mathbf{p}, \mathbf{Q}) = \min_{\mathbf{q}\in\mathbf{Q}} \|\mathbf{p} - T(\mathbf{q}; \hat{\mathbf{M}})\| \tag{6}$$

where $d(,)$ is a distance metric. Then the model will be re-estimated according to (5) after correspondence refinement, and so on. The iteration will be terminated when $d(,)$ is stable.

III. PROPOSED ALGORITHM

As mentioned before, area-based and feature-based image registration methods have their own strengths and limitations in the context of the ETDRS protocol. In this work, we propose a hybrid registration approach for ETDRS images,

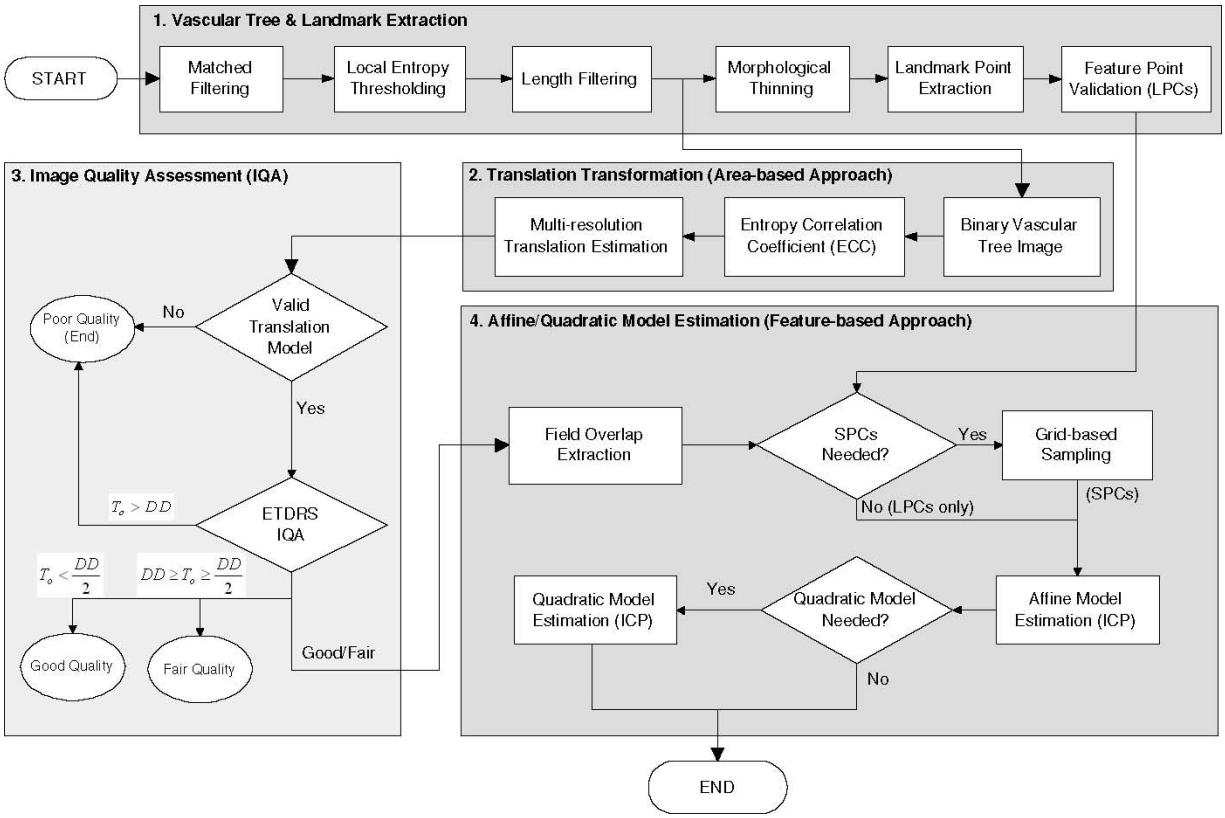


Fig. 2. Flowchart of the proposed algorithm. Landmark point correspondences (LPCs) and sampling point correspondences (SPCs).

which effectively takes the advantages of both area-based and feature-based methods in one flow. As shown in Fig. 2, the proposed algorithm is composed of three major steps, which are discussed in detail below.

A. Vascular Tree Extraction

Vascular tree extraction plays an important role in most retinal registration algorithms since the vascular tree is the most prominent anatomical structure in the retina. Vascular tree segmentation can be very useful for various purposes such as diagnosis and registration [28], [29], [39]–[47]. The vascular tree is considered to be the most appropriate representation for the image registration applications due to the three reasons: 1) it maps the whole retina; 2) it does not move except in very few diseases; and 3) it contains enough information for the localization of some anchor points [15]. Recently, we developed an efficient method to extract the vascular tree from retinal images which is composed of four steps [28], [29]. First, a match filter is applied to enhance the prominence of blood vessels [48]. Second, a local entropy-based thresholding scheme is used which takes into account the spatial distribution of gray levels and can well preserve the spatial structures in the binarized image [33]. Subsequently, a length filtering technique is used to remove misclassified pixels or insignificant small segments. For the match filtering results are shown in Fig. 4(a) and (b), the entropy-based thresholding with length filtering results are shown in Fig. 4(c) and (d) where we can see blood vessels are clearly segmented from the background. The binary vascular tree will be used for area-based registration. Then a morphological thinning operation is employed to obtain the centerline of the vascular tree. Finally, vascular crossover/bifurcation points, which will be used for feature-based registration, are located by a two-step window-based probing process. An initial probing process is first conducted to find potential branching points by locating the location where its 3×3 neighborhood contains more than three pixels belong to the thinned vascular tree. Then a 11×11 window is applied to all potential branching points to identify true branching points. If there are more than two vessel pixels of on the window boundary, it will be marked as a bifurcation/crossover point, as shown in Fig. 4(g) and (h).

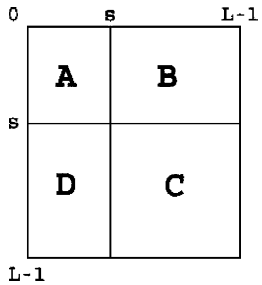


Fig. 3. Quadrants of cooccurrence matrix [33].

In the following, we briefly address the local-entropy thresholding technique, and more details can be found in [33]. Given the cooccurrence matrix of the image F is an $P \times Q$

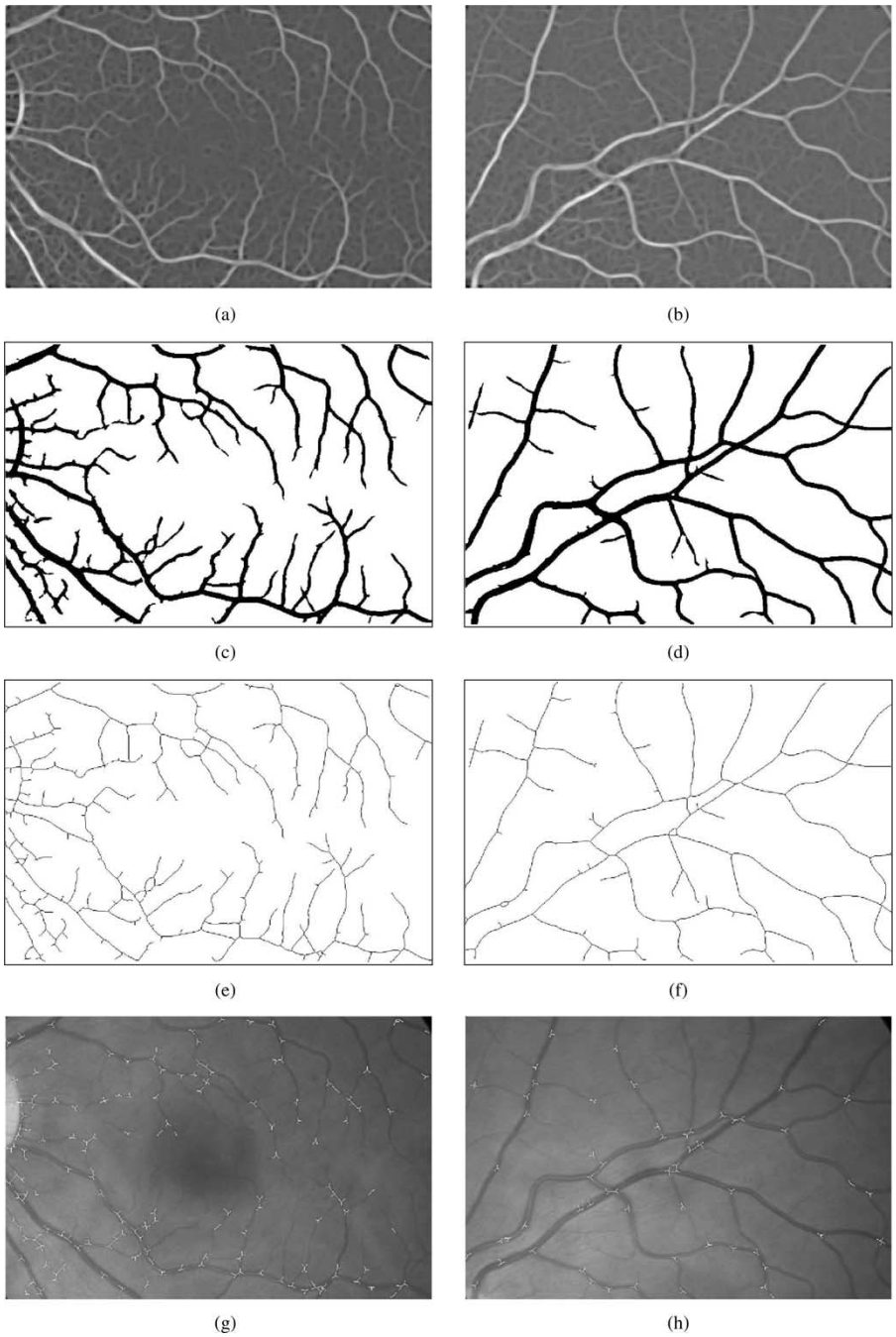


Fig. 4. Vascular tree and the centerline extraction of fields 2 and fields 4 of Julie's retinal images. (a) Match-filtered image: field 2. (b) Match-filtered image: field 4. (c) Binary image: field 2. (d) Binary image: field 4. (e) Thinned binary image: field 2. (f) Thinned binary image: field 4. (g) Crossover/bifurcation points: field 2. (h) Crossover/bifurcation points: field 4.

dimensional matrix $T = [t_{ij}]_{P\times Q}$ that gives an idea about the transition of intensities between adjacent pixels, indicating spatial structural information of an image. Depending upon the ways in which the gray level i follows gray level j, different definitions of cooccurrence matrix are possible. Here, we make the cooccurrence matrix asymmetric by considering the horizontally right and vertically lower transitions. In the following, the probability of cooccurrence of gray levels i and j can be denoted as p_{ij}. If s, $0 \le s \le L-1$, is a threshold, then s can partition the cooccurrence matrix into 4 quadrants, namely A, B, C, and D (Fig. 3).

Let us define the following quantities:

$$P_A = \sum_{i=0}^{s}\sum_{j=0}^{s} p_{ij}$$

$$P_C = \sum_{i=s+1}^{L-1}\sum_{j=s+1}^{L-1} p_{ij}. \tag{7}$$

By normalizing the probabilities within each individual quadrant, such that the sum of the probabilities of each quadrant equals one, we get the following cell probabilities for different quadrants:

$$P_{ij}^{A} = \frac{p_{ij}}{P_A} \text{ and } P_{ij}^{C} = \frac{p_{ij}}{P_C}. \tag{8}$$

Then the second-order entropy of the object can be defined as

$$H_A^{(2)}(s) = -\frac{1}{2}\sum_{i=0}^{s}\sum_{j=0}^{s} P_{ij}^{A}\log_2 P_{ij}^{A}. \tag{9}$$

Similarly, the second-order entropy of the background can be written as

$$H_C^{(2)}(s) = -\frac{1}{2}\sum_{i=s+1}^{L-1}\sum_{j=s+1}^{L-1} P_{ij}^{C}\log_2 P_{ij}^{C}. \tag{10}$$

The total second-order local entropy of the object and the background can be written as

$$H_T^{(2)}(s) = H_A^{(2)}(s) + H_C^{(2)}(s). \tag{11}$$

The gray level corresponding to the maximum of $H_T^{(2)}(s)$ gives the optimal threshold for object-background classification. Two vascular tree extraction examples and feature point identification are shown in Fig. 4.

B. Translation Estimation

The translation estimation is crucial for the hierarchical model estimation approach, since it will define a constraint for the higher-order model estimation as well as a restriction for correspondence selection. As seen from Fig. 4, manually determining a rough translation between fields $2/4$ in this image set is difficult because of small overlaps with complex vascular tree structures. In addition, not every pair of retinal images complies exactly with ETDRS field definition standard. Moreover, the point correspondences are not particularly distinguishable from each other. Instead of using crossover/bifurcation points, the translation estimation here is implemented by an area-based method which is based on a binary vascular tree due to the following three reasons: 1) the vascular tree is, undeniably, the most prominent structure and spans all ETDRS seven fields; 2) MI or ECC may not robust when the contrast/intensity distributions are not consistent within each image, and the binary image will greatly enhance the strength of MI and ECC computation; 3) in the ETDRS protocol, geometric distortion in seven-field images usually are not significant with negligible rotation and scaling, resulting in a relatively small searching space.

1) Binary ECC: Traditionally, MI or ECC based registration has been used on grayscale images. However, MI or ECC is not robust when the contrast/intensity distributions within each image field are not consistent, invalidating the statistical dependency across images. Therefore, we, instead, compute the ECC based on binary vascular tree images. Given two binary vascular trees images, $\mathbf{I}_u^{'}$, and $\mathbf{I}_v^{'}$, we estimate the translation model that aligns images $\mathbf{I}_u$ and $\mathbf{I}_v$ by maximizing the ECC between $\mathbf{I}_u^{'}$, and $\mathbf{I}_v^{'}$, as defined in the following:

$$\hat{\mathbf{M}} = arg\max_{\mathbf{M}} ECC(\mathbf{I}_u^{'}, T_0(\mathbf{I}_v^{'}; \mathbf{M})) \tag{12}$$

where $T_0(\mathbf{I}_v^{'}; \mathbf{M})$ translates image $\mathbf{I}_v^{'}$ with model $\mathbf{M}$. The ECC is defined in (4). Sample plots of binary ECC for every possible translation at the coarsest scale (downsampled by 16) are shown in Fig. 5(a) and (b) where the distinct peaks, indicating the optimal translations, can be easily identified. One might wonder, though, why not just employ the simple logical operation XOR. Here, we also tested XOR by developing an objective function defined as follows:

$$\hat{\mathbf{M}} = \arg\max_{\mathbf{M}} \frac{\text{XOR}(\mathbf{I}_u^{'}, T_0(\mathbf{I}_v^{'}; \mathbf{M})}{H \times W} \tag{13}$$

where $H \times W$ is the size of the overlaps. Example plots of XOR logical operation are shown in Fig. 5(c) and (d). Although XOR can also perform well in an image pair with insignificant distortions, XOR output does not give a distinguishable narrow peak as shown in Fig. 5(c). Moreover, in the case of retinal images with moderate geometric distortions, XOR does not provide the noticeable peaks, or it sometimes fails to locate the accurate translation as shown in Fig. 5(d).

We want to further manifest the advantages of ECC over XOR by defining an energy concentration function to numerically evaluate the peak quality of ECC's and XOR's outputs. The proposed energy concentration function is a simple indicator of peak distinction, which is defined as follows:

$$\Psi_K = \frac{\sum_{k=1}^{K}\phi_k^2}{\sum_{i=1}^{N_r}\sum_{j=1}^{N_c}(\psi_{i,j})^2} \times 100\% \tag{14}$$

where Ψ_K is the energy proportion of K highest peaks, ϕ_k is the kth largest peak value, $\psi_{i,j}$ is the ECC/XOR output under translation (i, j), and N_r and N_c are the numbers of all possible translations in row and column, respectively.

We compare ECC and XOR on 72 image pairs in terms of peak distinction by using Ψ_K as shown in Table III. It is shown that ECC is substantially better than XOR in estimating the optimal translation between a binary image pair. This might be due to the statistical capability of ECC which makes ECC's peaks much more distinct than those of XOR.

2) Binary Local Entropy: It is possible that there are multiple competitive peaks in ECC outputs, indicating several possible translations as shown in Fig. 6. Sometimes the highest peak does not necessarily represent the optimal translation due to the

TABLE III
ENERGY CONCENTRATION OF BINARY IMAGE-BASED ECC VERSUS LOGICAL OPERATION XOR

Fields	ECC: Avg. Ψ		XOR: Avg. Ψ	
	Ψ_1	Ψ_3	Ψ_1	Ψ_3
1/2	41.2%	71.5%	0.06%	0.17%
1/6	17.4%	27.6%	0.06%	0.17%
1/7	18.6%	29.4%	0.08%	0.23%
2/3	42.1%	58.6%	0.05%	0.15%
2/4	48.0%	72.8%	0.09%	0.25%
2/5	33.8%	67.8%	0.05%	0.15%
Avg. Ψ	40.6%	54.6%	0.07%	0.19%

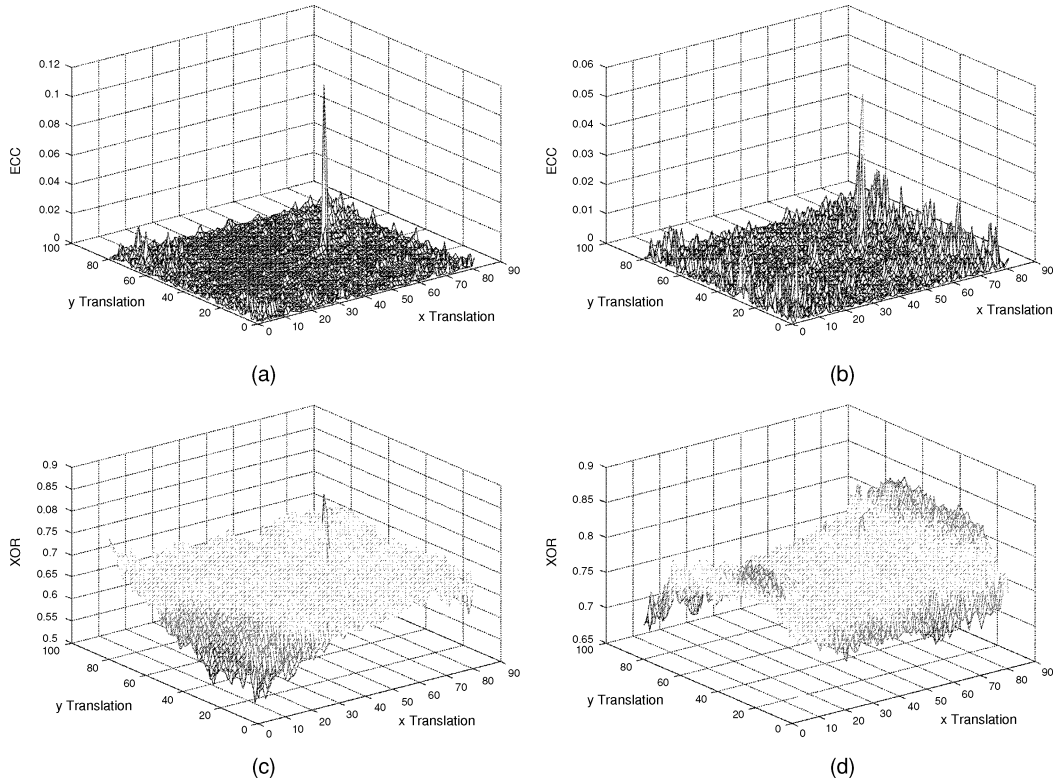


Fig. 5. Sample plots of binary image-based ECC versus logical operation XOR at every possible translation in the coarsest scale (downsampled by 16). (a) The binary image-based ECC of Julie's fields $1/2$; (b) The binary image-based ECC of Julie's fields $2/3$ (c) The logical operation XOR of Julie's fields $1/2$; (d) The logical operation XOR of Julie's fields $2/3$.

geometric distortion and the dissimilarity of binary image pairs. Therefore, an auxiliary criterion is needed to select the right peak. Since the local entropy can indicate the spatial structure in an image, we use it to evaluate resemblance between two overlaps. Let $\mathbf{I}^{(u)}$ and $\mathbf{I}^{(v)}$ represent the overlaps from binary vascular tree images $\mathbf{I}_u^{'}$ and $\mathbf{I}_v^{'}$, respectively. The unique optimal translation is obtained by

$$\hat{\mathbf{M}} = \arg\min_{\mathbf{M}\in\Omega} |H^{(2)}(\mathbf{I}^{(u)}) - H^{(2)}(T_0(\mathbf{I}^{(v)};\mathbf{M}))| \qquad (15)$$

where $\Omega = \{\mathbf{M}_1, \ldots, \mathbf{M}_N\}$ is a set of all possible translations with sufficiently large ECC values according to (12). $H^{(2)}$ is the local entropy defined in (2).

3) Multi-Resolution Searching Strategy: Given the translation model, there are only two parameters in the searching space which still could be very large due to the high-resolution nature of ETDRS images. Instead of going through every possible translation in the original scale, a multiresolution searching scheme is developed in order to reduce the computational complexity. A binary image is first represented in a pyramid of multiple resolutions from the coarsest scale to the finest scale. Let $\mathbf{I}_u^{'} = \{\mathbf{I}_u^{(0)}, \mathbf{I}_u^{(1)}, \mathbf{I}_u^{(2)}, \ldots, \mathbf{I}_u^{(J)}\}$ and $\mathbf{I}_v^{'} = \{\mathbf{I}_v^{(0)}, \mathbf{I}_v^{(1)}, \mathbf{I}_v^{(2)}, \ldots, \mathbf{I}_v^{(J)}\}$ represent the finest to the coarsest scales of binary images $\mathbf{I}_u^{'} = \mathbf{I}_u^{(0)}$ and $\mathbf{I}_v^{'} = \mathbf{I}_v^{(0)}$, respectively. First, the algorithm finds a binary-ECC peak, an optimal translation, at the coarsest scale, i.e., $\mathbf{I}_u^{(J)}$ and $\mathbf{I}_v^{(J)}, \mathbf{M}^{(J)} = \{(r_u^{(J)}, s_u^{(J)}), (r_v^{(J)}, s_v^{(J)})\}$ where $(r_u^{(J)}, s_u^{(J)})$ and $(r_v^{(J)}, s_v^{(J)})$ are two coordinates in two images showing the optimal translation between them. Then $\mathbf{M}^{(J)}$ can specify a constrained searching neighborhood at the finer scale, i.e., $\mathcal{N}(\mathbf{M}^{(J)}) = \{(2r_u^{(J)} + i, 2s_u^{(J)} + j), (2r_v^{(J)} + m, 2s_v^{(J)} + n) \,|\, i, j, m, n = -5, \ldots, 5\}$, where the optimal translation at scale $J-1$ scale can be obtained as follows:

$$\mathbf{M}^{(j-1)} = \arg\max_{\mathcal{N}(\mathbf{M}^{(J)})} \mathrm{ECC}\left(\mathbf{I}_u^{(j-1)}, \mathbf{I}_v^{(j-1)}\right) \qquad (16)$$

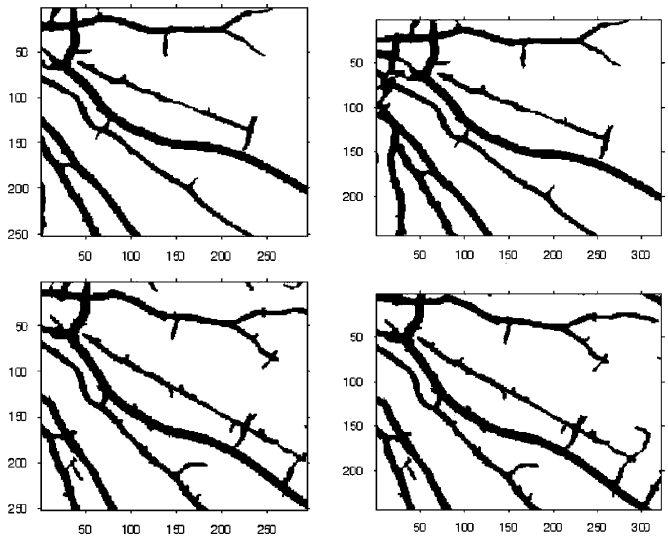


Fig. 6. Overlaps of two possible translation models (left column and right column) that produce similar values of ECC peaks for fields $1/7$.

where $j = J, J-1, \ldots, 1$. This procedure starts from the coarsest scale, and it is repeated until the finest scale is reached where an optimal translation, $\mathbf{M}^{(0)}$, is achieved at the pixel-level for images $\mathbf{I}_u^{(0)}$ and $\mathbf{I}_v^{(0)}$.

If the translation model $\mathbf{M}^{(0)}$ is reliable, we can move forward to higherorder models, such as affine/quadratic models, to obtain better registration performance. However, if $\mathbf{M}^{(0)}$ is incorrect, when there is significant geometric distortion between an image pair (unlikely to happen for ETDRS images) or there is no obvious vascular tree structure in overlaps (especially when image clarity is poor), the ECC-based translation estimation of binary images fails (this also indicates poor quality of ETDRS field coverage). In this work, we define two criteria based on the ECC output to check the correctness of $\mathbf{M}^{(0)}$. The first is energy concentration which measures the energy percentage of top three peaks as defined in (14), i.e., Ψ_3. The other is peak distinction that is the ratio of top two peaks, i.e., $\Phi = \phi_1/\phi_2$ where ϕ_k is the kth largest peak. A valid translation model usually leads to large Ψ_3 and Φ, indicating a sufficiently good match.

C. Affine/Quadratic Transformations

Usually, feature-based methods are supposed to be more reliable than area-based approaches if sufficient and accurate feature points are available and a proper translation model can be obtained which can greatly facilitate the subsequent higherorder transformation estimation. We here employ a feature-based scheme in order to refine the registration transformation from the zeroth-order to the higher-order model, where two different types of feature points, i.e., landmark points and sampling points, are involved, as illustrated in Fig. 2.

1) Landmark Point Correspondences (LPCs): Landmark points are the crossover/bifurcation points of vascular tree. The initial translation model achieved from Section III-B and landmark points obtained from Section III-A, are employed as the rudimentary guideline to establish the initial set of LPCs, C', for the first-order affine model

$$C' = \{(\mathbf{p}_i, \mathbf{q}_j) \,|\, \text{dist}(T(\mathbf{q}_j; \hat{\mathbf{M}}), \mathbf{p}_i) \leq \text{err}, \mathbf{p}_i \in \mathbf{P}, \mathbf{q}_j \in \mathbf{Q}\} \tag{17}$$

where $\mathbf{P}$ and $\mathbf{Q}$ are the sets of N_u and N_v landmark points from $\mathbf{I}_u^{'}$ and $\mathbf{I}_v^{'}$, respectively, $\hat{\mathbf{M}}$ is the initial translation model, dist$(,)$ denotes the Euclidean distance, and err is a threshold (e.g., 30 for the affine model and five for quadratic model). We need to have one-to-one matchings for all LPCs. However, there is no guarantee that LPCs in C' are one-to-one matching. In fact, a specific landmark point in $\mathbf{P}$ may have multiple matches, a single match, or no match at all in $\mathbf{Q}$. Therefore, we create a similarity matrix, $S = \{s_{i,j} \,|\, i = 1, \ldots, N_u; j = 1, \ldots, N_v\}$, with the purpose of assuring one-to-one matching for every LPC. The similarity measure, $s_{i,j}$ as defined below, is a coarse measure to quantify the resemblance between $\mathbf{p}_i$ and $\mathbf{q}_j$

$$s_{i,j} = \begin{cases} \mathbf{x}_i \cdot \mathbf{y}_j, & (\mathbf{p}_i, \mathbf{q}_j) \in C' \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

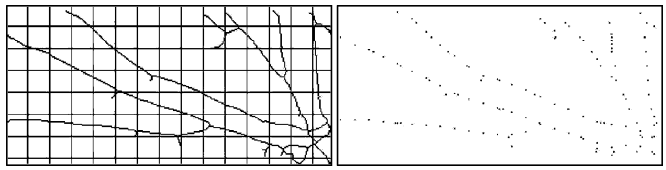


Fig. 7. Example of the sampling process. (Left) A grid is placed on the thinned binary vascular tree. (Right) The sampling points.

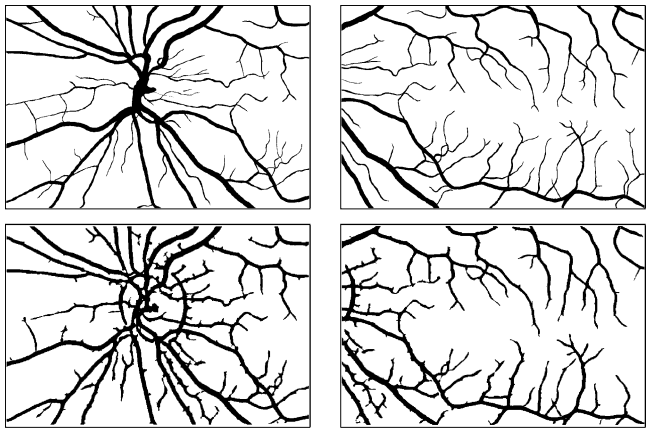
where $\mathbf{x}_i$ and $\mathbf{y}_j$ are obtained by placing a 9×9 window centered at $\mathbf{p}_i$ and $\mathbf{q}_j$ on the thinned images of $\mathbf{I}_u^{'}$ and $\mathbf{I}_v^{'}$, respectively. One-to-one LPC matchings are achieved by

$$C = \{(\mathbf{p}_i, \mathbf{q}_j) \,|\, j = \arg \max_{j \in 1, \ldots, N_v} s_{i,j}, \quad i = 1, \ldots, N_u.\}. \tag{19}$$

After C is obtained, we need to examine the reliability of LPCs. Let σ_x^2 and σ_y^2 be the second central moments of vertical/horiztonal coordinates of LPCs in the overlap of size $H_o \times W_o$. We define $(H_o)/(\sigma_x), (W_o)/(\sigma_y)$ to show how the LPCs spread in the overlap. If they are large (e.g., >4), LPCs are likely to cluster together in a small area. Then the sampling process is needed to involve more feature points for image registration.

2) Sampling Point Correspondences (SPCs): Sampling points can be acquired by imposing grid lines on the thinned vascular tree, where the intersections between blood vessels and the grid are marked as sampling points as shown in Fig. 7. If LPCs are not sufficient to estimate affine (at least 3) or quadratic (at least 6) models, SPCs are introduced to facilitate feature-based registration by providing some auxiliary information to LPCs. However, SPCs are usually less trustable compared with LPCs, since they are acquired from the thinned vascular tree which often exhibits strong linearity and are likely be linearly dependent if the vascular tree is sparse. Therefore, SPCs are only involved when LPCs do not meet certain criteria, and more details are discussed in the simulation. Given a set of sampling points, SPCs can be achieved in the same way as LPCs defined in (17) and (19). It is worth noting that the similarity metric defined in (19) is less effective for the SPCs where the vessel exhibits strong straightness. Thus SPCs are more valuable when the vascular tree has strong non-linearity, indicating more LPCs.

3) Iterative Closest Point (ICP): Both LPCs and SPCs are the input for the ICP algorithm which is a procedure to refine the model estimation by finding the closest point $\mathbf{q} \in \mathbf{Q}$ for every $\mathbf{p} \in \mathbf{P}$ given transformation $\mathbf{M}$. During the ICP iteration, *bad* LPCs/SPCs, i.e., the ones with significant Euclidean distances after the transformation (e.g., six pixels), are eliminated. The affine model is reestimated at each iteration by finding the minimum mean square error solution according to (5). The iteration is terminated when the model is stable or there is no change to LPCs/SPCs. We will proceed to the quadratic model if the registration error of the affine model is significant and we have at least Six LPCs. According to [13], a quadratic model is a good approximation to retinal surface since retinal surface approximates a sphere. The quadratic model could be useful when the overlap is significant. However,



(a) (b)

Fig. 8. Hand-labeled groundtruths (top) and the extracted vascular trees using the proposed algorithm (bottom) of (a) field 1 and (b) field 2.

in the case of ETDRS, we prefer not to proceed to the quadratic model when overlaps between two images are very limited or LPCs are not sufficient (say less than Six). It may not be robust to estimate the transformation of the whole image based only on small overlaps where the spatial information across different fields is too limited. The higher order model, sometimes, could introduce significant distortions to the image pairs with small overlaps.

IV. Experimental Analysis

The ETDRS images were provided by Inoveon[2] which is a medical services company delivering solutions to diagnose eye-related diseases, mainly diabetic retinopathy. Since the ETDRS protocol defines seven stereoscopic fields in each eye (two sets of six pairs for each eye), there will 24 image pairs for one patient. Totally, 504 pairs collected from 21 patients were involved for algorithm evaluation. These images were taken with a Kodak DCS520 digital camera coupled with a Zeiss FF450 fundus camera which has the original resolution of 1152×1728. In order to make the size of registered images tractable, we down-sampled all images to 600×900. Still the final registration results are very large, nearly 2000×3000.

A. Vascular Tree Extraction

We hand-labeled vascular tree structures for several retinal images for the comparison purpose, as shown in Fig. 8. Performance of blood vessel extraction is evaluated by using the true positive and the false positive rates as in [39]. Any pixel which was hand-labeled as vessel and the algorithm labeled as vessel is counted as true positive. Any pixel which was hand-labeled as nonvessel and the algorithm labeled as vessel is counted as false positive. The true positive rate is calculated by normalizing true positive by the total pixel number of the hand-labeled vessel. The false positive rate is calculated by normalizing false positive by the total pixel number of the hand-labeled non-vessel.

[2]http://www.inoveon.com/index.html

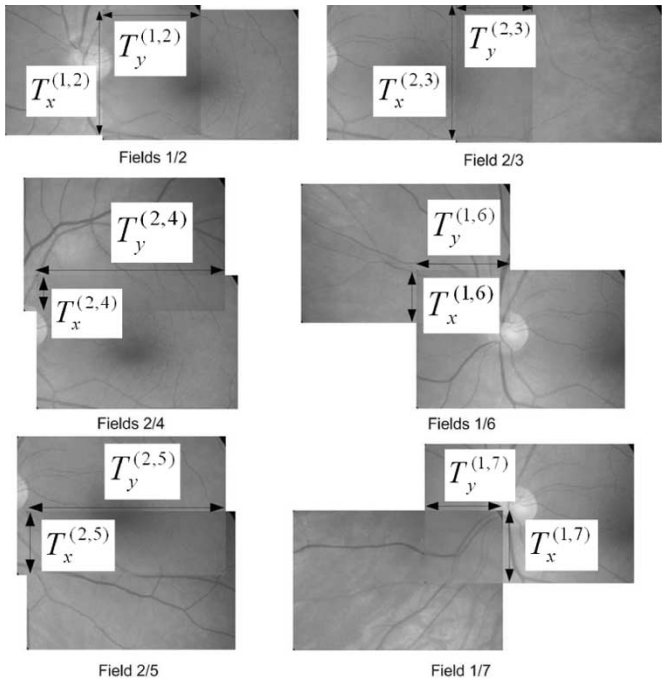


Fig. 9. Vertical/horizontal displacements of six ETDRS image pairs.

TABLE IV
Ideal Vertical/Horizontal Displacements

Field Pair	Desired vertical/horizontal displacements ($T_x^{'}/T_y^{'}$) $H = 600$ and $W = 900$ are the image size.	
	$T_x^{'}$	$T_y^{'}$
1/2	H	$0.5W$
1/6	$0.5H - 0.5DD$	$0.5W - 0.5DD$
1/7	$0.5H - 0.5DD$	$0.5W - 0.5DD$
2/3	H	$0.5W$
2/4	$0.5H - 0.5DD$	W
2/5	$0.5H - 0.5DD$	W

Are shown in Fig. 8, the true positive rates for fields 1 and 2 are 0.9050 and 0.8854, and the false positive rates for fields 1 and 2 are 0.0798 and 0.0619. In general, our vascular segmentation results are thicker than ground-truths, and this thickening effect is introduced by the match filter which tends to increase the width of vessels. Due to the symmetric property of the match filter, this artifact has negligible effect on the area-based translation estimation. Moreover, the feature-based transformation estimation mainly relies on thinned centerlines of vascular tree which are relatively stable after match filtering.

B. IQA Simulation

We assume that photographers can indicate whether the positions of optic disc and macular in fields 1 and 2 are accurate and they also can obtain the diameter of optic disc (DD). In order to perform IQA, we define vertical/horizontal displacements for each ETDRS field pair in Fig. 9, and Table IV lists all ideal translation displacements for each field pair according to the ETDRS field definition specified in Table I and Fig. 1. Then the IQA of field coverage can be easily implemented by computing the offset, i.e., T_o, which is the difference between

TABLE V
REGISTRATION ERROR AND OVERLAPS BETWEEN FIELDS

Fields	Median Overlap	Median Registration Errors (pixels)					
		Translation	LPCs-Only Algorithm		LPCs-SPCs Algorithm		Proposed Algorithm
			Affine	Quadratic	Affine	Quadratic	
1/2	52.97%	$19.21^{(84,82)}$	$3.15^{(78,76)}$	$2.36^{(78,76)}$	$2.86^{(82,78)}$	$2.09^{(82,78)}$	$2.15^{(82,78)}$
1/6	18.91%	$27.78^{(84,68)}$	$2.39^{(47,43)}$	$1.42^{(34,32)}$	$2.33^{(69,66)}$	$1.87^{(69,58)}$	$1.91^{(69,66)}$
1/7	22.50%	$22.58^{(84,72)}$	$2.55^{(53,49)}$	$1.75^{(32,30)}$	$2.33^{(73,68)}$	$2.09^{(73,61)}$	$1.97^{(73,68)}$
2/3	44.24%	$20.80^{(84,78)}$	$3.64^{(68,63)}$	$2.95^{(68,66)}$	$3.43^{(79,74)}$	$2.41^{(79,72)}$	$2.62^{(79,74)}$
2/4	25.39%	$19.10^{(84,84)}$	$1.97^{(76,75)}$	$1.08^{(62,59)}$	$1.96^{(84,81)}$	$1.19^{(84,80)}$	$1.79^{(84,81)}$
2/5	32.17%	$18.38^{(84,81)}$	$2.43^{(67,66)}$	$1.96^{(62,59)}$	$2.66^{(81,78)}$	$2.02^{(81,74)}$	$2.11^{(81,78)}$
Average	32.70%	$21.30^{(504,465)}$	$2.68^{(389,372)}$	$1.92^{(336,322)}$	$2.59^{(468,443)}$	$1.94^{(468,423)}$	$2.04^{(468,445)}$

Totally, there are 84 retinal image pair for each field pair. Some pairs might be unable to perform registration without the sampling process. (N_1, N_2) : N_1 indicates the number of image pairs are registered. N_2 specifies the number of registered pairs after manual validation.

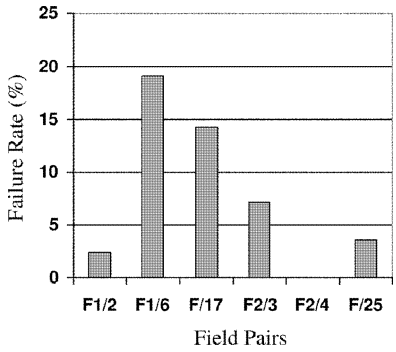


Fig. 10. Failure rates of translation estimation for six ETDRS field pairs.

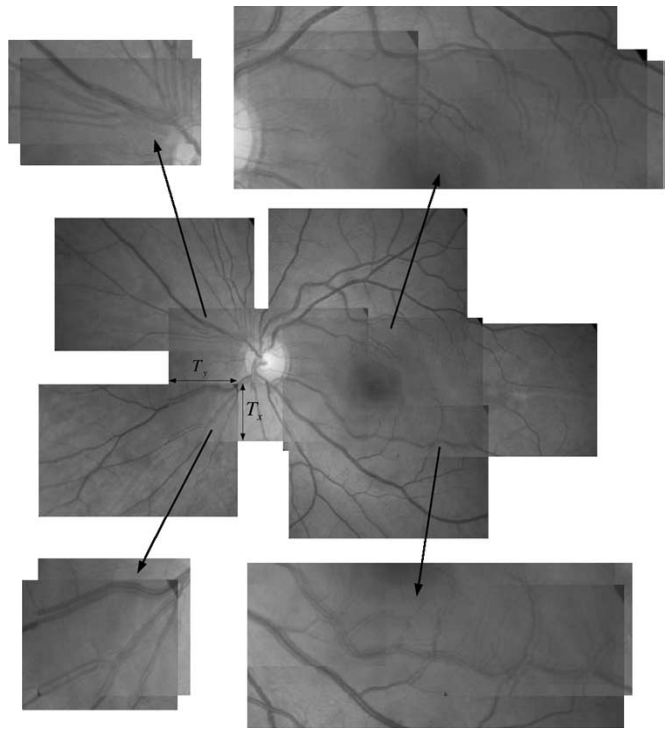


Fig. 11. Translation model-based registration result where the vertical/horizontal displacements in fields $1/7$ are depicted. The registration errors are as follows: fields $1/2 = 15.81$, fields $1/6 = 22.06$, fields $1/7 = 13.83$, fields $2/3 = 12.60$, fields $2/4 = 21.56$, and fields $2/5 = 18.97$. The median error is 17.39.

the ideal displacements and the actual ones given by the translation model. By comparing T_o with DD, an image pair can be classified as "good," "fair," or "poor" quality. For example, all six pairs in Fig. 9 are "good" pairs. Since a correct translation model guarantees the IQA validity, the IQA accuracy is mainly determined by the success rate of translation estimation. Actually, the actual IQA accuracy could be higher, since most cases when the ECC-based translation estimation fails are due to the significant incompliance to the ETDRS field definition which are "poor" pairs.

C. Registration Performance Evaluation

In this section, we present both quantitative and qualitative analysis of the three major techniques on 504 ETDRS image pairs. It was found that the ICP iteration runs 1–5 rounds in most cases. We use the registration error defined in (5) to evaluate the algorithm performance. The numerical results are summarized in Table V. Given the binary vascular tree, the total computational time for image registration per pair is approximately 20 s on a 2.8 GHz PC and the Matlab 6.5 platform without algorithm optimization.

1) Translation Model Estimation: The registration error of the translation model over the whole data set is 21.30 pixels, as shown in Table V. This performance is still acceptable for the IQA purpose, since DD is usually near 200 pixels. Fig. 11 shows an example of the translation-based registration result, where ghost lines due to registration errors are clearly visible as seen in the zoomed-in regions. Out of 504 registered pairs, there are 39 pairs are rejected by manual validation, and the success rate is 92.3%. As mentioned before, we have defined two criteria to determine the credibility of the translation model based on the ECC output, i.e., energy concentration of top three peaks (Ψ_3) and peak distinction of top two peaks (Φ). From those bad pairs, we found both Ψ_3 and Φ are relatively small compared with others. Thus we use $\Psi_3 > 13\%$ or $\Phi > 2.0$ as two empirical conditions to accept the estimated translation model. These two criteria are able to identify 36 bad pairs out of 39 pairs manually rejected. The failure rate of translation estimation is shown in Fig. 10, where the failure rates of fields $1/6$ and $1/7$ are relatively higher due to small overlaps (usually less than 20%).

2) Affine/Quadratic Model Estimation: We have 468 image pairs for which the translation model is accepted and a

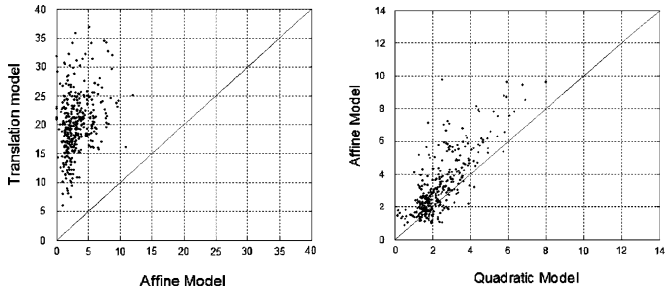


Fig. 12. Error comparisons for consecutive transformation models.

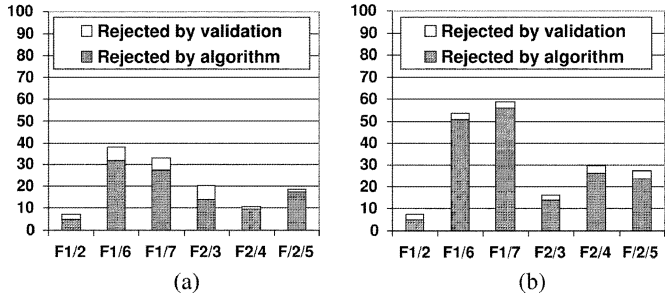


Fig. 13. Failure rates of 468 image pairs that proceed from the translation model to affine and quadratic models based on LPCs only.

higher-order model is applied. Fig. 12 shows the change of registration errors resulting from consecutive models.

It is clearly seen that the error is dramatically reduced from a lower order model to a higherorder model. From Table V, the error of the affine model is 2.68 pixels which is much less than that of the translation model, and the error is further reduced to 1.92 for the quadratic model. The numerical values demonstrate that the hierarchical strategy can improve the robustness and accuracy of image registration progressively [13]. However, since the affine and quadratic models need at least three and six LPCs, respectively, many pairs cannot proceed to high-order models due to insufficient LPCs, and also there are some registered pairs are rejected by manual validation, the failure rates are shown in Fig. 13.

3) LPCs and SPCs: Although affine/quadratic models can greatly reduce the registration error, they have high failure rates due to the lack of sufficient LPCs in many field pairs, especially when using the quadratic model in fields $1/6$ and $1/7$ where few LPCs exist. The use of SPCs could reduce failure rates by adding more feature points for feature-based registration. As shown in Table V, after involving SPCs, all 468 pairs proceed to affine/quadratic models, among which 443 pairs and 423 pairs are manually validated for the affine and quadratic models, respectively. The overall failure rates are shown in Fig. 14. Moreover, SPCs slightly reduce for the registration error of affine model (from 2.68 to 2.59). It is interesting to note that SPCs significantly improve the accuracy of the quadratic model in fields $1/2$ and $2/3$ where LPCs are usually sufficient, and SPCs have some negative effect in other field pairs where LPCs are often limited. This is because SPCs are less reliable and they can only provide some auxiliary information to LPCs for image registration.

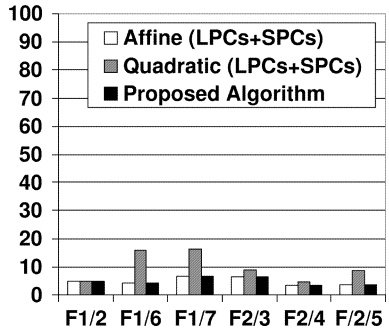


Fig. 14. Failure rates (after manual validation) of affine/quadratic models (using both LPCs and SPCs) and the proposed algorithm for six field pairs.

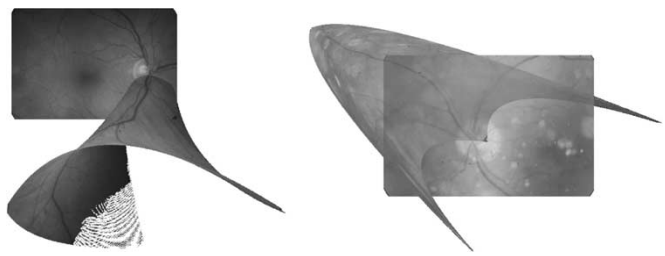


Fig. 15. Two examples of using the quadratic transformation on fields $2/5$ and fields $1/6$ with insufficient LPCs, where the registration errors are 0.4 and 1.4.

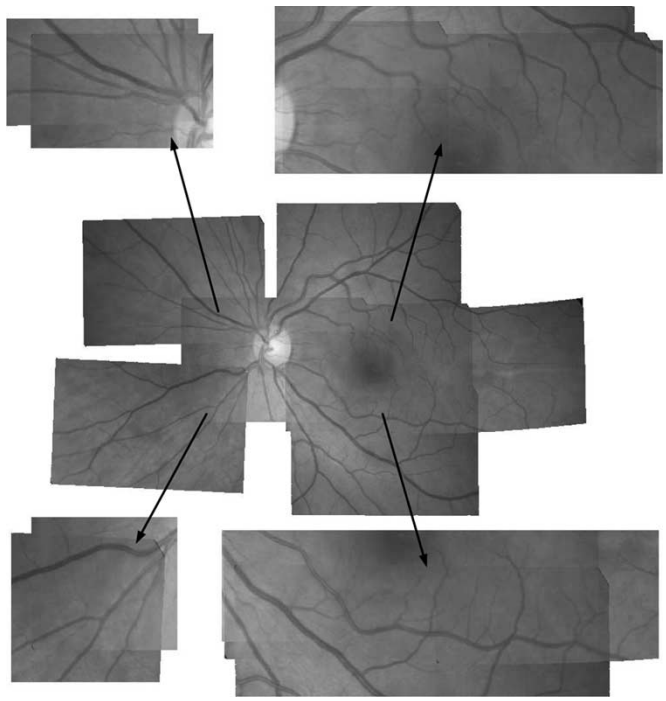


Fig. 16. Final registration result of an ETDRS image set. The quadratic transformation is applied to fields $2/3$. The affine model is applied to fields $1/2, 1/6, 1/7, 2/4$, and $2/5$. The registration errors are given as as follows: fields $1/2 : 2.00$, fields $1/6 : 0.52$, fields $1/7 : 1.23$, fields $2/3 : 1.66$, fields $2/4 : 2.78$, and fields $2/5 : 1.49$. The median error is 1.58. The sampling points (SPCs) are involved in fields $1/6, 2/3$ and fields $2/4$.

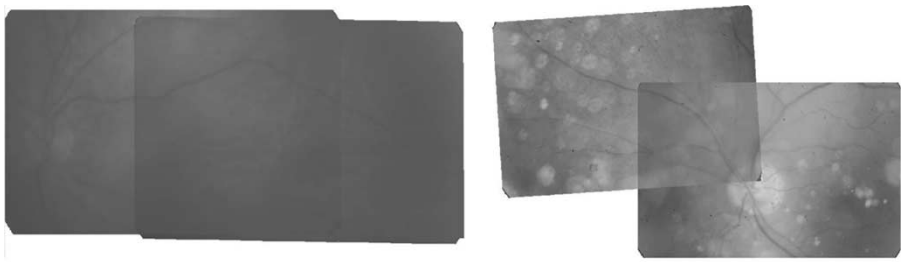


Fig. 17. (a) Example of a poor quality retinal image pair with no LPC in the overlap. (b) Example of a retinal image pair with pathology.

D. Discussions of the Proposed Algorithm

The above techniques constitute the major components of the proposed algorithm. It is worth mentioning that the registration error may not be always trustable to evaluate the registration performance, especially when LPCs are limited and SPCs are involved for higher-order model estimation. Fig. 15 shows two examples where the registration errors fails to indicate the accuracy of high-order registration results. Therefore, we have to effectively combine above techniques into one flow where the proposed algorithm is able to: 1) adaptively select an appropriate transformation model; 2) determine whether SPCs have to be involved, and more importantly; and 3) reject invalid registered pairs. As mentioned before, we have developed a set of if-then conditions based on the simulation results which allow the algorithm to achieve these objectives. We summarize these conditions as follows:

- *Condition 1) Translation Model*: The translation model is essential to the whole algorithm, and its correctness is validated by the ECC output. If energy concentration and peak distinction are sufficiently large, i.e., $\Psi_3 > 13\%$ or $\Phi > 2.0$, the translation model is accepted based on which the IQA is performed, and the registration proceeds to higher-order models. Otherwise, the algorithm terminates. This condition can reject 36 pairs out of 39 pairs where the translation estimation fails due to poor image quality and significant geometric distortion.
- *Condition 2) Affine/Quadratic Models*: For ETDRS images, the affine model is usually sufficient in most field pairs, and going to the quadratic model may be risky or could be wrong due to small overlaps. No matter whether or not SPCs are involved, the quadratic model is only applied when there are at least 6 LPCs. It is found that most pairs undergoing the quadratic model are in the fields $1/2$ and $2/3$.
- *Condition 3) LPCs and SPCs*: There are two cases under which SPCs have to be involved: 1) when the number of LPCs is less than 3; 2) or if LPCs cluster together in a small area, i.e., $(H_o)/(\sigma_x) > 4.0$ or $(W_o)/(\sigma_y) > 4.0$. We found that the sparsely distributed LPCs usually lead to undesired registration results, even though the reported registration errors are small. LPCs are expected to be uniformly distributed in the overlaps, otherwise SPCs will be involved.

Above empirical conditions are incorporated into the proposed algorithm, as shown in Fig. 2, which is able to provide a high success rate and low registration errors, as shown in Fig. 14 and Table V. On the on hand, the success rate of the proposed algorithm (445 out of 468, i.e., 95.1%) is higher than that of the affine model without/with sampling ($372/443$ out of 468, i.e., $79.5\%/94.7\%$). On the other hand, the registration error (2.04) is lower than that the affine model without/with sampling $(2.68/2.59)$, and it is close the that of the quadratic model without/with sampling $(1.92/1.94)$. One registration example of the proposed algorithm is shown in Fig. 16, where the quadratic model is applied in fields $2/3$ and the SPCs are involved in fields $1/6, 2/3$, and $2/4$. This six field pairs also pass the IQA with the good quality of ETDRS field coverage according to Table IV.

Moreover, the proposed algorithm is able to work well for retinal image pairs that are defocused/blur with very poor contrast as well as image pairs with pathologies, as shown in Fig. 17. In Fig. 17(a), blood vessels are barely perceptible and there is no LPC. In Fig. 17(b), spot lesions are proliferated across both images. The proposed algorithm still can successfully registers these image pairs. A single area-based or feature-based approach alone may not be sufficient for ETDRS image registration. The proposed hybrid registration algorithm shows the promising performance on ETDRS images.

Overall speaking, there are a few major characteristics of the proposed algorithm. 1) It can handle retinal images with relatively small overlaps, and the performance is less dependent on the size of overlaps due to the power of ECC estimation on binary vascular tree images. 2) No LPC is required for the algorithm. 3) The algorithm selects an appropriate model for each field pair and involves SPCs when it is necessary. However, the proposed algorithm may fail when the geometric distortion between image pairs is significant (unlikely to happen in the ETDRS case), where the ECC-based translation estimation may be incorrect.

V. Conclusion

This paper presents an ETDRS retinal image registration algorithm that effectively combines both area-based and feature-based methods into one flow. Three empirical conditions are used to: 1) select an appropriate transformation model; 2) determine whether the sampling process is needed; and 3) reject invalid registration results, so that we can maximize the success

rate and minimize the registration error. The proposed method can be used for the IQA purpose in terms of ETDRS field definition and to facilitate the implementation of ETDRS protocols in clinical trials.

ACKNOWLEDGMENT

The authors would like to thank the three referees whose comments and suggestions have greatly improved this paper.

REFERENCES

[1] L. G. Brown, "A survey of image registration rechniques," *ACM Comput. Surv.*, vol. 24, pp. 325–376, Dec. 1992.

[2] B. Zitova and J. Flusser, "Image registration methods: A survey," *Image Vision Comput.*, vol. 21, pp. 977–1000, 2003.

[3] P. A. van den Elsen, E. D. Pol, and M. A. Viergever, "Medical image matching—A review with classification," *IEEE Eng. Med. Biol. Mag.*, vol. 12, no. 1, pp. 26–39, Mar. 1993.

[4] J. B. A. Maintz and M. A. Viegever, "A survey of medical image registration," in *Medical Image Analysis*. Oxford, U.K.: Oxford Univ. Press, 1998.

[5] H. Lester and S. Arridge, "A survey of hierarchical nonlinear medical image registration," *Pattern Recogn.*, vol. 32, no. 1, pp. 129–149, 1999.

[6] N. Mouravliansky, G. K. Matsopoulos, K. Delibasis, and K. S. Nikita, "Automatic retinal registration using global optimization techniques," in *Proc. Int. Conf. Eng. Med. Biol. Society*, vol. 2, Nov. 1998, pp. 567–570.

[7] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita, "Automatic retinal image registration scheme using global optimization techniques," *IEEE Trans. Inf. Technol. Biomed.*, vol. 3, no. 1, pp. 47–60, Mar. 1999.

[8] N. Ritter, R. Owens, J. Cooper, R. H. Eikelboom, and P. P. V. Saarloos, "Registration of stereo and temporal images of the retina," *IEEE Trans. Med. Imag.*, vol. 18, no. 5, pp. 404–418, May 1999.

[9] D. Lloret, J. Serrat, A. M. Lopez, A. Soler, and J. J. Villaneuva, "Retinal image registration using creases as anatomical landmarks," in *IEEE Proc. Int. Conf. Pattern Recognition*, vol. 3, Sep. 2000, pp. 203–206.

[10] A. V. Cideciyan, "Registration of ocular fundus images: An algorithm using cross-correlation of triple invariant image descriptors," *IEEE Eng. Med. Biol. Mag.*, vol. 14, no. 1, pp. 52–58, Jan.–Feb. 1995.

[11] L. Ballerini, "Temporal matched filters for integration of ocular fundus images," in *Proc. IEEE Conf. Digital Signal Processing*, vol. 2, Jul. 1997, pp. 1161–1164.

[12] M. Skokan, A. Skoupy, and J. Jan, "Registration of multimodal images of retina," in *Proc. IEEE Conf. Eng. Med. Biol.*, vol. 2, Oct. 2002, pp. 1094–1096.

[13] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, "A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 3, pp. 347–364, Mar. 2002.

[14] ——, "A feature-based technique for joint, linear estimation of higher-order image-to-mosaic transformations: Mosaicing the curved human retina," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 3, pp. 412–419, Mar. 2002.

[15] F. Zana and J. Klein, "A multimodal registration algorithm of eye fundus images using vessels detection and hough transform," *IEEE Trans. Biomed. Eng.*, vol. 18, no. 5, pp. 419–428, May 1999.

[16] ——, "A registration algorithm of eye fundus images using a bayesian hough transform," in *IEEE Int. Conf. Image Processing Proc. Applications*, vol. 2, Jul. 1999, pp. 479–483.

[17] H. Shen, C. V. Stewart, B. Roysam, G. Lin, and H. L. Tanenbaum, "Frame-rate spatial referencing based on invariant indexing and alignment with application to online retinal image registration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 3, pp. 379–384, Mar. 2003.

[18] C. V. Stewart, C. L. Tsai, and B. Roysam, "The dual-bootstrap iterative closest point algorithm with application to retinal image registration," *IEEE Trans. Med. Imag.*, vol. 22, no. 11, pp. 1379–1394, Nov. 2003.

[19] F. Laliberte, L. Gagnon, and Y. Sheng, "Registration and fusion of retinal images—An evaluation study," *IEEE Trans. Med. Imag.*, vol. 22, no. 5, pp. 661–673, May 2003.

[20] C. L. Tsai, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, "Covariance-driven retinal image registration initialized from small sets of landmark correspondences," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Jul. 2002, pp. 333–336.

[21] E. H. Zhang, Y. Zhang, and T. X. Zhang, "Automatic retinal image registration based on blood vessels feature point," in *Proc. IEEE Int. Conf. Machine Learning Cybernetics*, vol. 4, Nov. 2002, pp. 2010–2015.

[22] W. E. Hart and M. H. Goldbaum, "Registering retinal images using automatically selected control point pairs," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Nov. 1994, pp. 576–580.

[23] C. Heneghan, P. Maguire, N. Ryan, and P. de Chazal, "Retinal image registration using control points," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Jul. 2002, pp. 349–352.

[24] J. Park, J. M. Keller, P. D. Gader, and R. A. Schuchard, "Hough-based registration of retinal images," in *Proc. IEEE Int. Conf. Systems, Man, Cybernetics*, vol. 5, Oct. 1998, pp. 4550–4555.

[25] P. Jasiobedzki, "Registration of retinal images using adaptive adjacency graphs," in *Proc. IEEE Symp. Comput. Based Med. Syst.*, Jun. 1993, pp. 40–45.

[26] A. M. Mendonca, A. Campilho, and J. M. R. Nunes, "A new similarity criterion for retinal image registration," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Sep. 1994, pp. 696–700.

[27] U. N. I. O. H. N. National Eye Institute Clinical studies database: Early Treatment diabetic retinopathy study (ETDRS). [Online]. Available: Http://www.nei.nih.gov/neitrials/viewStudyWeb.aspx?id=53

[28] T. Chanwimaluang and G. Fan, "An efficient blood vessel detection algorithm for retinal images using local entropy thresholding," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 5, May 2003, pp. V21–V24.

[29] ——, "An efficient algorithm for extraction of anatomical structures in retinal images," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, Sep. 2003, pp. I1093–I1096.

[30] S. R. Fransen, T. C. Leonard-Martin, W. J. Feuer, and P. L. Hildebrand, "Clinical evaluation of patients with diabetic retinopathy: Accuracy of the inoveon diabetic retinopathy-3dt system," *Amer. Acad. Ophthalmol.*, vol. 109, no. 3, pp. 595–601, Mar. 2002.

[31] C. J. Rudnisky, B. J. Hinz, M. T. S. Tennant, A. R. de Leon, and M. D. J. Greve, "High-resolution stereoscopic digital fundus photography versus contact lens biomicroscopy for the detection of clinically significant macular edema," *Amer. Acad. Ophthalmol.*, vol. 109, no. 2, pp. 267–274, Feb. 2002.

[32] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: The University of Illinois Press, 1949.

[33] N. R. Pal and S. K. Pal, "Entropic thresholding," *Signal Process.*, vol. 16, pp. 97–108, 1989.

[34] J. P. W. Pluim, J. B. Antoine, and M. A. Viegever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.

[35] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recogn.*, vol. 32, no. 1, pp. 71–86, 1999.

[36] A. Collignon, "Multi-modality medical image registration by maximization of mutual information," Ph.D. dissertation, Catholic Univ., Leuven, Belgium, 1998.

[37] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multi-modality medical image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 4, pp. 187–198, Apr. 1997.

[38] P. J. Besl and N. D. McKay, "A method for registration of 3-D shape," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.

[39] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, Mar. 2000.

[40] A. Can, H. Shen, J. N. Turner, H. L. Tanenbaum, and B. Roysam, "Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms," *IEEE Trans. Inf. Technol. Biomed.*, vol. 3, no. 2, pp. 125–138, Jun. 1999.

[41] X. Jiang and D. Mojon, "Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal

images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 1, pp. 131–137, Jan. 2003.

[42] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.

[43] A. Pinz, S. Bernogger, P. Datlinger, and A. Kruger, "Mapping the human retina," *IEEE Trans. Med. Imag.*, vol. 17, no. 4, pp. 606–619, Aug. 1998.

[44] C. Sinthanayothin, J. F. Boyce, H. L. Cook, and T. H. Williamson, "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," *B. F. Ophthalmol.*, pp. 902–910, Feb. 1999.

[45] W. Tan, Y. Wang, and S. Lee, "Retinal blood vessel detection using frequency analysis and local-mean interpolation filters," in *Proc. SPIE Medical Imaging: Image Processing*, vol. 4322, 2001 , pp. 1373–1384.

[46] N. H. Solouma, A. M. Youssef, Y. A. Badr, and Y. M. Kadah, "A new real-time retinal tracking systems for image-guided laser treatment," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 9, pp. 1059–1067, Sep. 2002.

[47] G. Lin, C. V. Stewart, B. Roysam, K. Fritzsche, G. Yang, and H. L. Tanenbaum, "Predictive scheduling algorithms for real-time feature extraction and spatial referencing: Application to retinal image sequences," *IEEE Trans. Biomed. Imag.*, vol. 51, no. 1, pp. 115–125, Jan. 2004.

[48] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Trans. Med. Imag.*, vol. 8, no. 3, pp. 263–269, Sep. 1989.

Thitiporn Chanwimaluang (S'02) received the B.Eng. degree in electrical engineering from Chulalongkorn University, Bangkok, Thailand, and the M.S. degree in electrical engineering from Pennsylvania State University, University Park, in 1996 and 2001, respectively. She is currently pursuing the Ph.D. degree in the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater.

She was with Siemens, Thailand, from 1996 to 1997 and then she was with Thai Telephone and Telecommunications, Thailand, from 1997 to 1998. Her recent research interests include image registration and 3-D reconstruction.

Guoliang Fan (S'97–M'01–SM'05) received the B.S. degree in automation engineering from Xi'an University of Technology, Xi'an, China, the M.S. degree in computer engineering from Xidian University, Xi'an, China, and the Ph.D. degree in electrical engineering from University of Delaware, Newark, in 1993, 1996, and 2001, respectively.

From 1996 to 1998, he was a Graduate Assistant in the Department of Electronic Engineering, Chinese University of Hong Kong. Since 2001, he has been with the School of Electrical and Computer Engineering, Oklahoma State University (OSU), Stillwater, where he is currently an Assistant Professor. His research interests are signal/image/video processing, biomedical imaging and remote sensing applications.

Dr. Fan was awarded the First Prize in the 1997 IEEE Hong Kong Section Postgraduate Student Paper Contest and the First Prize in the 1997 IEEE Region 10 (Asia-Pacific) Postgraduate Paper Contest. He is a recipient of the National Science Foundation (NSF) CAREER award (2004). He received the Halliburton Excellent Young Teacher Award from OSU in 2004.

Stephen R. Fransen received the B.Sc. degree in chemistry and M.D. degree from the University of Manitoba, Canada, in 1979 and 1985, respectively, and completed an Ophthalmology Residency at the University of Oklahoma (OU), Norman, in 1989.

In 1990, he joined the faculty of the OU Department of Ophthalmology, where he led the design of the Internet-based diabetic retinopathy evaluation system now licensed exclusively to Inoveon. He is a tenured Associate Professor of Ophthalmology at OU and continues to devote a portion of his time to the care of patients with retinal diseases. He is a founder and the Chief Clinical Officer of Inoveon Corporation, Oklahoma City.

Dr. Fransen received the Dean A. McGee Eye Institute and a Retina Fellowship with J. Donald M. Gass, at the University of Miami, Bascom Palmer Eye Institute, in 1990.